

David Feller
Environmental Molecular Sciences Laboratory,
Pacific Northwest National Laboratory, MS K1-96
P.O. Box 999, Richland, Washington 99352

ABSTRACT

A variety of molecular properties, including atomization energies, bond lengths and harmonic vibrational frequencies were computed for more than 80 experimentally well-characterized molecules, many of which were taken from the Gaussian-2 and G2/97 collections. This body of data was stored in the Environmental Molecular Sciences Laboratory Computational Results Database and was subsequently analyzed to determine the sensitivity of each property towards the electron correlation treatment and the quality of the underlying Gaussian basis set. In light of the importance of the basis set in limiting the ultimate accuracy of the theoretical predictions, a wide range of correlation consistent basis sets (including, in some cases, up through aug-cc-pV6Z) was used. These were combined with five popular levels of theory, ranging from simple Hartree-Fock theory up through coupled cluster theory. The importance of core/valence, scalar relativistic, atomic spin-orbital effects and more extensive correlation recovery were also examined in terms of their impact on agreement with experiment. All of these effects can contribute on the order of 1 - 2 kcal/mol or more to quantities such as a heat of formation.

I. INTRODUCTION

The accuracy of five widely-used electronic structure methods in reproducing experimental atomization energies (ΣD_e), electron affinities, proton affinities, ionization potentials, vibrational frequencies and geometries were recently examined by Feller and Peterson.¹ Methods included Hartree-Fock (HF) theory, second- and fourth-order Møller-Plesset perturbation theory (MP2 and MP4), coupled cluster theory with single and double excitations (CCSD) and coupled cluster theory with a quasiperturbative treatment of triples (CCSD(T)). In addition to the raw errors resulting from each specific (method/ basis set) pair, an attempt was made to assess a given method's *intrinsic* error, i.e. the deviation with respect to experiment in the limit of a complete 1-particle basis set.

Throughout the development of quantum chemistry, up until the early 1990s, the primary sources of error in most electronic structure calculations were more-or-less evenly attributable to relatively poor quality basis sets and an inadequate treatment of the electron correlation problem. However, following the development of the correlation consistent basis sets²⁻⁵ it became possible to exploit the regularity of these basis sets to perform an effective extrapolation to the complete basis set (CBS) limit. This was accomplished by fitting results obtained from successively larger and larger basis sets with one of a number of simple functional forms. Feller and Peterson examined three extrapolations. The first of these is the empirically motivated exponential form, given for total energies by the expression:

$$E(x) = E_{\text{CBS}} + b e^{-cx} \quad (1)$$

where x is an index associated with each basis set, $x=2$ (DZ), 3(TZ), 4(QZ), etc.⁶⁻¹⁰ The second is a combined Gaussian/exponential function:^{11,12}

$$E(x) = E_{\text{CBS}} + b e^{-(x-1)} + c e^{-(x-1)^2} \quad (2)$$

And, the final expression is:

$$E(\ell_{\text{max}}) = E_{\text{CBS}} + b/(\ell_{\text{max}} + 1)^4 \quad (2)$$

where ℓ_{max} is the maximum angular momentum present in the basis set.¹³ For second and third row correlation consistent basis sets, x (eq. 1) = ℓ_{max} (eq. 2).

II. PROCEDURE

Most geometries were optimized with a gradient convergence criterion of $1.5 \times 10^{-5} E_h/a_0$, corresponding to the "tight" criterion defined in Gaussian-94,¹⁴ using the same level of theory as was used in evaluating the desired property. For example, CCSD(T) thermochemical properties were evaluated at the optimal CCSD(T) geometries. A looser convergence criterion of $1.0 \times 10^{-4} E_h/a_0$ was necessary for the largest basis set CCSD(T) optimizations.

Unless otherwise noted, open shell energies were based on unrestricted Hartree-Fock (UHF) zeroth order wavefunctions and were performed with the Gaussian-94 program.¹⁴ Orbital symmetry and equivalence restrictions were not imposed in atomic calculations. Closed shell CCSD(T) calculations were performed with MOLPRO-97¹⁵ and Gaussian-94. CCSDT calculations were obtained from ACES II.¹⁶ All calculations were performed on a 16 processor Silicon Graphics, Inc. PowerChallenge, a 32 processor SGI Origin 2000 or an SGI/Cray J90 at D.O.E.'s National Energy Research Supercomputing Center. The largest CCSD(T) calculation reported in this study included 734 functions. As in the previous study, results were stored and analyzed using the Environmental Molecular Sciences Laboratory (EMSL) Computational Results Database,¹⁷ which currently contains over 32,000 entries.

Atomization energies were corrected for the effects of core/valence (CV) correlation (which is not included in normal frozen core calculations run with programs like Gaussian) by performing all electron calculations with the cc-pCVQZ basis sets, which are specially designed for this purpose. The 1s pairs of electrons for third period elements was treated as frozen cores.

Atomic spin-orbit and molecular/atomic scalar relativistic corrections were also appended to our nonrelativistic atomization energies and are denoted ΔE_{SO} and ΔE_{SR} , respectively. The former account

for the improper description of the atomic asymptotes, since atomic energies determined by our calculations correspond to an average over the possible spin multiplets. In some cases, such as the $^2\Pi$ states of molecules like CH and OH, there is an additional *molecular* spin-orbit correction due to the splitting of the $^2\Pi_1$ and $^2\Pi_j$ states. Spin-orbit corrections were taken from the atomic and molecular values reported by Dunning and coworkers,^{10,12} which are based on the experimental values of Herzberg¹⁸ and Moore.¹⁹ Scalar relativistic corrections were obtained from configuration interaction wavefunctions including single and double excitations (CISD) using the cc-pVTZ basis set. The CISD(FC) wavefunction was used to evaluate the dominant 1-electron Darwin and mass-velocity terms in the Breit-Pauli Hamiltonian.

Our results will be compared to experimental atomization energies extrapolated to 0 K, both with and without zero point energies (ZPEs), i.e. $\Sigma D_0(0K)$ and $\Sigma D_e(0K)$, respectively. A majority of the experimental data used in this report was taken from the NIST-JANAF Tables (4th. Edition)²⁰ and Huber and Herberg.²¹ Our zero point energies were based on harmonic CCSD(T) frequencies.

III. RESULTS

Table 1 contains representative comparisons between CCSD(T) values of ΣD_0 and the corresponding, best available experimental values. The three right-most columns contain the differences with respect to experiment for the three CBS estimates, eqs. 1 - 3. For most of the 80 molecules the convergence in the frozen core binding energies is slow enough that fairly large basis sets are required. Of all the methods tested, only CCSD(T) shows continuous improvement in the level of agreement with experiment as the basis set size increases. If a double zeta basis set is largest that can be afforded, MP2 actually gives better statistical agreement with experiment.

In general, the errors arising from the use of $ZPE = 1\Sigma\omega$, where ω are harmonic frequencies obtained from CCSD(T) calculations, is relatively small. The mean absolute deviation with respect to experiment, ϵ_{MAD} , for frozen core CCSD(T) is ~ 1.5 kcal/mol, with worse case errors as large as 4.5 kcal/mol. This value includes adjustments for atomic spin-orbit effects. Core/valence corrections range from essentially zero to as much as 7.1 kcal/mol, and can be of either sign. By including core/valence corrections ϵ_{MAD} drops into the 0.7 - 0.8 kcal/mol range. Relativistic corrections produce no *overall* change in ϵ_{MAD} , but did reduce the maximum errors. Although scalar relativistic corrections tend to decrease the binding energy, they can be of either sign and vary from near zero to 2.4 kcal/mol.

A correction should also be applied for the difference between CCSD(T) and full CI. Very little is known about this difference because CCSD(T) recovers such a large percentage of the correlation energy that determining energies with even greater accuracy is a very difficult task. We recently examined the impact of higher order correlation effects on the dissociation energies of HF, N₂ and CO. Among the higher order methods examined were two variations of coupled cluster theory (CCSDT and CCSD(TQ)) and two approximations to full configuration interaction. Again, basis sets were chosen from the correlation consistent family of basis sets, with the largest being the aug-cc-pVQZ set. Polarized valence double zeta quality basis sets were found to yield corrections that differed substantially from larger basis set results. At the double zeta level, higher order corrections increased the binding energies, whereas calculations with triple and quadruple zeta basis sets gave the opposite effect. Although the absolute magnitude of the higher order corrections was small for these diatomics, they were nonetheless significant in light of a target accuracy of ± 1 kcal/mol. Among molecules composed of first-through-third period elements, such as those in the G2 and G2/97 collections, the contribution to ΣD_0 from higher order correlation effects could easily exceed 1 kcal/mol. CCSD(TQ) often overestimated the higher order correction, sometimes exceeding the estimated full configuration interaction result by a factor of three.

Normally, because CCSD(T) is based on only a single reference configuration, it would not be expected to describe transition states very well. However, for some transition states, where the Hartree-Fock configuration still constitutes a significant component in the transition state wavefunction, CCSD(T) may perform as well as explicit multireference methods like configuration interaction (CI). The transition state for the reaction $H_2CO \rightarrow H_2 + CO$ has been obtained with a wide assortment of theoretical methods. Compared to complete active space CI, CCSD(T) does a good job of predicting the barrier height and is far less costly.

VII. CONCLUSIONS

CCSD(T) atomization energies were computed for a set of 80 molecules with reliable experimental data. Basis sets were taken from the augmented correlation consistent family and represent some of the largest Gaussian basis sets currently available. By using any of the three complete basis set extrapolations (eqns. 1 - 3), it is possible to enter a regime where core/valence, scalar relativistic, atomic spin-orbit, anharmonic ZPE effects or higher order correlation effects can become as important as the remaining error due to the use of finite basis set. Failure to account for any one of these effects can lead to errors on the order of several kcal/mol or more in particularly troublesome cases. On the other hand, in fortunate cases some of these corrections can nearly cancel. The overall mean absolute deviation is below 1 kcal/mol.

Although CCSD(T) suffers from some of the same limitations as other single-reference methods, it is currently the most accurate *ab initio* electronic structure technique that can be applied with large basis sets to small molecules. Coupled cluster theory *without* the inclusion of triple excitations was found to be frequently less accurate for atomization energies than second order perturbation theory. CCSD(T) is not capable of describing large regions of most potential energy surfaces, but nonetheless it may do quite well for certain transition states.

ACKNOWLEDGMENTS

This research was supported by the U. S. Department of Energy under Contract No. DE-AC06-76RLO 1830. The Pacific Northwest National Laboratory is a multiprogram national laboratory operated by Battelle Memorial Institute.

REFERENCES

- (1) Feller, D.; Peterson, K. A. *J. Chem. Phys.* **1998**, *108*, 154.
- (2) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.
- (3) Woon, D. E.; Peterson, K. A.; Dunning, T. H., Jr. **to be published**.
- (4) Wilson, A. K.; Mourik, T. v.; Dunning, T. H., Jr. *J. Mol. Struct. (Theochem)* **1996**, *388*, 339.
- (5) van Mourik, T.; Dunning, T. H., Jr. *Theor. Chem. Acc.* **to be published**.
- (6) Feller, D. *J. Chem. Phys.* **1992**, *96*, 6104.
- (7) Xantheas, S. S.; Dunning, T. H., Jr. *J. Phys. Chem.* **1993**, *97*, 18.
- (8) Feller, D. *J. Chem. Phys.* **1993**, *98*, 7059.
- (9) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1993**, *99*, 1914.
- (10) Peterson, K. A.; Kendall, R. A.; Dunning, T. H., Jr. *J. Chem. Phys.* **1993**, *99*, 1930.
- (11) Peterson, K. A.; Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1994**, *100*, 7410.
- (12) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1994**, *101*, 8877.
- (13) Martin, J. M. L. *Chem. Phys. Lett.* **1996**, *259*, 679.
- (14) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andreas, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*, E.3, Gaussian, Inc. Pittsburgh PA, 1996.
- (15) Werner, H. J.; Knowles, P. J.; Almlöf, J.; Amos, R. D.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Elbert, S. T.; Hampel, C.; Lindh, R.; Lloyd, A. W.; Meyer, W.; Nicklass, A.; Peterson, K. A.; Pitzer, R. M.; Stone, A. J.; Taylor, P. R.; Mura, M. E.; Pulay, P.; Schütz, M.; Stoll, H.; Thorsteinsson, T. *MOLPRO*, Universität Stuttgart, Stuttgart, Germany, University of Sussex, Falmer, Brighton, England, 1997.
- (16) Stanton, J. F.; Gauss, J.; Watts, J. D.; Nooijen, M.; Oliphant, N.; Perera, S. A.; Szalay, P. G.; Lauderdale, W. J.; Gwaltney, S. R.; Beck, S.; Balkova, A.; Bernholdt, D. E.; Baeck, K.-K.; Rozyczko, P.; Sekino, H.; Hober, C.; Bartlett, R. J.; Almlöf, J.; Taylor, P. R.; Helgaker, T.; Jensen, H. J. A.; Jørgensen, P.; Olsen, J. *ACES II*, University of Florida, 1998.
- (17) Feller, D. *J. Comp. Chem.* **1996**, *17*, 1571.
- (18) Herzberg, G. *Molecular Spectra and Molecular Structure I. Spectra of Diatomic Molecules*; Van Nostrand Reinhold Co., Inc.: New York, 1950.
- (19) Moore, C. E. *Atomic Energy Levels* Washington, D.C., 1949; Vol. U.S. National Bureau of Standards Circular 467, NBS.
- (20) Chase, M. W., Jr. *NIST-JANAF Tables (4th Edition)*, *J. Phys. Chem. Ref. Data*, **1998**, *Mono. 9*, Suppl. 1.
- (21) Huber, K. P.; Herzberg, G. *Molecular Spectra and Molecular Structure: Constants of Diatomic Molecules*; Van Nostrand Reinhold Co. Inc.: New York, 1979; Vol. 4.

Table 1. Contributions to CCSD(T) Atomization Energies (kcal/mol) for a Selected Subset of the G2 Molecules.^a

Molecule	ZPE Expt.			Atom. Expt. ΣD_0			Expt. ΣD_0	Erro r	w.r.t. Mixed	Expt. ℓ_{\max}
	ΣD_e	$1\Sigma_0$	ZPE	ΔE_{CV}	ΔE_{SR}	ΔE_{SO}	(0K)			
CH ₄ ($1A_1$)	418.9	27.8	27.6	1.3	-0.2	-0.1	392.5 ± 0.1	-0.4	-0.4	-0.3
H ₂ O ($1A_1$)	232.5	13.5	13.3	0.5	-0.3	-0.2	219.35 ± 0.01	-0.4	-0.6	-0.4
HF ($1\Sigma^+$)	141.4	5.9	5.9	0.2	-0.2	-0.4	135.2 ± 0.2	-0.1	-0.1	-0.1
SiH ₂ ($1A_1$)	153.6	7.3	7.3	0.0	-0.2	-0.4	144.4 ± 0.7	1.3	1.3	1.5
HCl ($1\Sigma^+$)	107.0	4.3	4.2	0.3	-0.2	-0.8	102.24 ± 0.5	-0.2	0.0	0.1
CO ($1\Sigma^+$)	258.2	3.1	3.1	0.9	-0.2	-0.3	256.2 ± 0.2	-0.7	-0.7	-0.5
N ₂ ($1\Sigma_g^+$)	226.1	3.4	3.4	1.0	-0.1	0.0	225.1 ± 0.4	-1.5	-1.1	-0.8
SO ₂ ($1A_1$)	257.6	3.9	4.4	1.0	-0.9	-1.0	254.0 ± 0.2	-1.2	-0.3	0.4
Cl ₂ ($1\Sigma_g^+$)	58.8	0.8	0.8	0.2	+0.7	-1.7	57.18 ± 0.01	0.0	0.4	0.7

^aThe atomic asymptotes were described with the UCCSD(T) method. The column labeled " ΣD_e " contains complete basis set estimates based on aug-cc-pVTZ through aug-cc-pV6Z basis sets. Theoretical zero point energies were obtained from CCSD(T) calculations. Errors for the exponential, mixed and $1/\ell_{\max}$ CBS atomization energies (ΣD_0), which appear in the three right-most columns, were computed as the difference between the theoretical value, defined as: $E(\text{CCSD(T)}(\text{FC})/\text{CBS}) - 12\nu_1 + \text{CV} + \text{scalar relativistic} + \text{atomic/molecular S.O.}$ and the bolded experimental value. For diatomics with a nonzero molecular spin-orbit contribution, e.g. OH ($^2\Pi$), the sum of the atomic and molecular contributions is included in ΔE_{SO} .